Particle Physics Data Grid Applications

David Malon

Argonne National Laboratory

malon@anl.gov

What is PPDG?

- **A Next-Generation Internet (NGI) project to build a wide-area data access infrastructure for high energy and nuclear physics experiments based upon emerging network and middleware technologies**

Goals

- **Design, develop, and deploy a network and middleware infrastructure capable of supporting data analysis and data flow patterns common to the many particle physics experiments represented by the participants**

Goals

- **Adapt experiment-specific software to operate in this wide-area environment and to exploit this infrastructure**

- **Instantiate and deliver an operating infrastructure for distributed data analysis by participating physics experiments**

Participants

- **Argonne National Laboratory**

- **Brookhaven National Laboratory**

- **California Institute of Technology**

- **Fermi National Accelerator Laboratory**

- **Lawrence Berkeley National Laboratory**

- **Stanford Linear Accelerator Center**

- **Thomas Jefferson National Accelerator Facility**

- **San Diego Supercomputer Center**
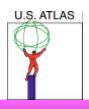
- **University of Wisconsin**

ATLAS PPDG Participants

- **Argonne National Laboratory**
  - u **David Malon (PPDG Applications Working Group convenor and ATLAS applications coordinator), Ed May, David Gunter, Larry Price; initial ATLAS data source (tile testbeam)**

- **Brookhaven National Laboratory**
  - u **Bruce Gibbard, Torre Wenaus; ATLAS data node**

- **Lawrence Berkeley National Laboratory**
  - u **Stu Loken (LBNL ATLAS rep); ATLAS data node**

CMS PPDG Participants

- **California Institute of Technology**

  - u **Harvey Newman (project co-PI), Julian Bunn (CMS applications coordinator); CMS data source**

- **Fermi National Accelerator Laboratory**

  - u **Vicky White, Ruth Pordes, Jim Amundson, Don Petravick, Igor Terkekhov (FNAL CMS reps); CMS data node**

First-year PPDG Services

- **High-Speed Site-to-Site File Replication Service**

- **Multi-Site Cached File Access Service**
  - u **based on deployment of file replica cataloging, transparent cache management, and data movement middleware**

Use of NGI Services

- **Differentiated services**

- **Distributed Caching**

- **Matchmaking and resource coscheduling**

ATLAS/CMS Joint Efforts

- **ATLAS & CMS are providing the "challenge problems" for PPDG--the applications that require that all the disparate components of the architecture work in concert**

- **Have articulated shared use cases to be addressed with a common architecture and common components, though both the data and the applications that consume them differ**

DOE/NSF ATLAS/CMS Review
January 2000

ATLAS/CMS Collaboration

- **Shared use cases** include simultaneous high speed file replication from data source to regional centers using differentiated services, and multiple concurrent analysis jobs requiring data that are replicated at several sites, including "local" regional centers

  - u **both ATLAS and CMS have challenge applications in which the files are Objectivity/DB databases**

ATLAS/CMS Collaboration

- **Common components include**

  - u **Globus middleware**

  - u **matchmaking from the University of Wisconsin**

  - u **cache management based in part upon HENP Grand Challenge components from LBNL/ANL**

  - u **storage management based on SRB from SDSC (but with interfaces that could be supported by FNAL-developed storage management software)**

Status

- **Originally a three-year project, now funded as a one-year effort (no FY 00 Congressional appropriation for NGI)**

    u **access to network testbeds will still be provided**

- **Will nonetheless provide the spadework for grid-based data access and computing for high energy and nuclear physics (see related presentations on GriPhyN and APOGEE)**

- **Scaled back deliverables include high speed file replication and location-transparent wide-area access to replicated data**